W690–W692 *Nucleic Acids Research, 2005, Vol. 33, Web Server issue*
doi:10.1093/nar/gki445

RibEx: a web server for locating riboswitches and other conserved bacterial regulatory elements

Cei Abreu-Goodger and Enrique Merino*

Departamento de Microbiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, 62210 Morelos, México

Received February 14, 2005; Revised and Accepted March 30, 2005

ABSTRACT

We present RibEx (riboswitch explorer), a web server capable of searching any sequence for known riboswitches as well as other predicted, but highly conserved, bacterial regulatory elements. It allows the visual inspection of the identified motifs in relation to attenuators and open reading frames (ORFs). Any of the ORF's or regulatory elements' sequence can be obtained with a click and submitted to NCBI's BLAST. Alternatively, the genome context of all other genes regulated by the same element can be explored with our genome context tool (GeConT). RibEx is available at http://www.ibt.unam.mx/biocomputo/ribex.html.

INTRODUCTION

Ribonucleic acids have become fashionable lately. Apart from their fundamental participation in transcription and translation, RNAs are clearly some of the most functionally diverse molecules in the cell. Recently, non-translated regions of several mRNAs have been found to be capable of regulating their own expression by binding specific metabolites with high affinity in complete absence of proteins [(1), reviewed in (2)]. These regulatory elements, termed riboswitches, appear to be highly conserved, the extreme case being that of the thiamine pyrophosphate (TPP) riboswitch, which has been found in all three kingdoms of life (3). Riboswitches comprise two parts, a sensing element or aptamer, which forms a complex structure capable of binding the metabolite, and an effector element, or expression platform capable of transforming the signal into a biological response. The aptamer is the most conserved, having been selected to bind an unchanging molecule such as a vitamin or an amino acid. Upon binding, a shift between two mutually exclusive RNA secondary structures in the effector element occurs. These pairs of structures of the expression platform can represent a transcriptional terminator/anti-terminator, a Shine-Dalgarno sequester/anti-sequester or even an active/inactive ribozyme (2,4). It is not uncommon for different organisms to use the same sensing element, yet different effector elements.

FINDING RIBOSWITCHES

Although the usual method to define a riboswitch involves locating a conserved secondary structure in the RNA molecule, the highly restricted nature of the sensing element argues that sequence alone should be enough to locate riboswitches correctly. We have previously developed a computer algorithm capable of finding bacterial regulatory motifs, based exclusively on sequence conservation in the regulatory regions of orthologous groups of genes (5). The main restrictions of our method are that a regulatory element must be closely associated with at least one COG (cluster of orthologous groups of proteins) (6) and it must be present in at least five non-redundant genomes. On the other hand, the advantage is that it is an automatic process, requiring no previous regulatory information to produce relevant results, and as such, can be easily run every time that new genomes or annotations are available.

We updated our previous results (5), taking into account 223 complete genomes. From these, a reduced set of 145 non-redundant organisms was obtained using CVtree (7). We were able to recover 10 out of the 11 currently reported riboswitches. Additionally, our results included many regulatory elements that are also known to depend on structured RNA for recognition, such as the Gram-positive T-box and the PyrR protein binding site. We thus call our set of regulatory elements: riboswitch-like elements (RLEs), given the fact that almost all the identified conserved signals were RNA-dependant regulatory elements.

RibEx is a web server that allows any user to easily find any RLE in the sequence of his/her interest. Since most known riboswitches are associated with attenuators, we have included the option of searching for transcriptional and translational attenuators, which can help in selecting the most likely candidates, as has been shown by Barrick *et al*. (4). Additionally, our web server displays representative drawings of the open

*To whom correspondence should be addressed. Tel: +52 777 329 16 29; Fax: +52 777 317 23 88; Email: merino@ibt.unam.mx

reading frames (ORFs) and their corresponding regulatory elements, any of which can be selected, in order to acquire its sequence for submission to NCBI's BLAST server (8). Every RLE is linked to a list of genes that are predicted to be subject to its regulation. The genome context of these genes, analyzed with our local GeConT web server (9), in addition to the scores of the pre-computed RLEs, can be of great assistance when evaluating the likelihood of a new prediction.

A great resource when working with RNA families is the Rfam database (10). We have used their models to annotate our RLEs. As of version 7.0, Rfam contains a total of 503 families, 125 of them are non-coding, and 11 of these are annotated as riboswitches. We were able to recover automatically all but one of these riboswitches, missing the *ykoK* element. Our matrices for the most abundant riboswitches perform very well when compared with the co-variance models used by Rfam (~90% coverage when analyzing bacterial sequences). Less common riboswitches (e.g. lysine and purine) are more difficult to model with sequence-based weight-matrices. Our method thus tends to recover between 70 and 80% of these Rfam members. Our data set also contains six more RLEs that coincide with an Rfam *cis*-regulating member and 341 RLEs that do not have a match and thus remain as predicted elements. We have calculated a *P*-value, assuming a hyper-geometrical distribution, for each RLE to be over-represented in a given COG or KEGG pathway (11). Thus, we provide every RLE with a tentative functional assignation.

As far as we know there are only two servers, beside ours, that can be used to locate riboswitches in a given sequence: riboswitch finder (12) which, in its current implementation, only searches for the purine-sensing riboswitch, and Rfam, that has an option to locate riboswitches in any sequence, but as co-variance searches have high computational requirements, the sequence length is limited to 2 kb. RibEx, in addition to performing searches on larger sequences, allows the user a greater view of the regulatory potential of his sequence, by showing the ORFs and predicted attenuators. The 341 predicted RLEs also make RibEx a great complement to the curated families contained in Rfam.

THE WEB SERVER

The server is divided into modules, which are written in, and tied together with Perl. A brief description of each module follows:

Riboswitch-like elements. The program takes the sequence provided and splits it into overlapping windows of 500 nt. Each of these smaller sequences are searched for the selected RLEs with MAST (13), using matrices obtained as detailed in our previous work (5). Our method defines each RLE as several non-overlapping motifs, so we restrict the search to 500 nt to avoid false positives where the individual motifs are too far apart. When an RLE passes the selected *E*-value cutoff, the positions, size of each motif and final score of the regulatory element are recorded.

Open reading frames. ORFs are predicted, as is commonly done for bacterial genomes. The default options are for a resulting protein of at least 80 amino acids beginning with a start codon (ATG, GTG or TTG) and ending with a stop codon (TAA, TAG or TGA). By default, fully overlapped ORFs are not shown.

Attenuators. These are predicted according to an algorithm developed in our group and described elsewhere (14). The predicted secondary structure of each attenuator and its free energy is recorded. Upon clicking on the image of the attenuator, an additional window will be opened showing this information. To avoid false positives, attenuators are only searched for in the region preceding each predicted ORF.

Web output. The web page is generated 'on the fly' by a Perl script that controls all the other modules. The images are generated using the GD graphics library, and the interactivity between windows and frames is provided with Javascript.

AN EXAMPLE

Figure 1 shows a typical RibEx output. The input sequence was a region of 4000 nt from around the *thiC* gene of *Bacillus cereus ATCC14579*. Immediately upstream from one of the ORFs (drawn as blue arrows) the three motifs that comprise the TPP riboswitch (red boxes) can be seen, as well as

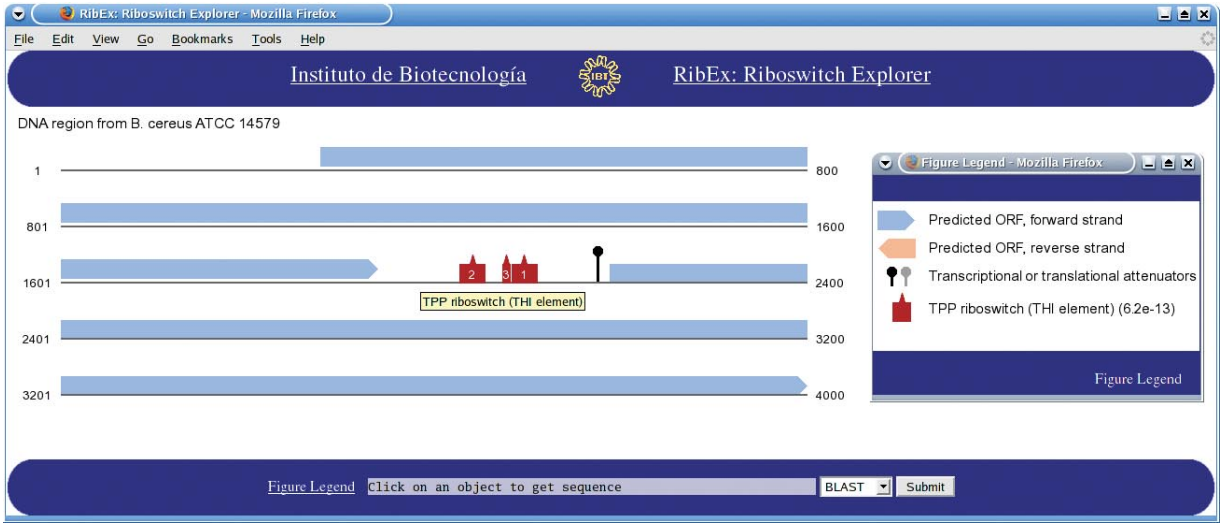

Figure 1. RibEx locates a thiamine riboswitch.

a transcriptional attenuator (black lollipop). A separate window acts as a figure legend indicating the score for each regulatory element found (in this case, only the TPP riboswitch). A typical scenario might include clicking on the second ORF, and sending the sequence to the BLAST web server, showing it to be identical to several ThiC proteins. Clicking on the TPP riboswitch motif in the figure legend box opens a window with the genes that are predicted to be regulated by this riboswitch, where the user can see how the motifs are distributed in different genomes. Taken together, and strengthened by the presence of a transcriptional attenuator, the user would have no trouble at all concluding that his sequence contains a bona fide riboswitch.

ACKNOWLEDGEMENTS

We wish to thank Ricardo Ciria for support in setting up the web server. This work was supported by CONACyT grant 44213-Q to E.M. and C.A.G. was supported by fellowships from CONACyT and DGEP-UNAM. The Open Access publication charges for this article were waived by Oxford University Press.

Conflict of interest statement. None declared.

REFERENCES

1. Winkler,W., Nahvi,A. and Breaker,R.R. (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*, **419**, 952–956.
2. Nudler,E. and Mironov,A.S. (2004) The riboswitch control of bacterial metabolism. *Trends Biochem. Sci.*, **29**, 11–17.
3. Sudarsan,N., Barrick,J.E. and Breaker,R.R. (2003) Metabolite-binding RNA domains are present in the genes of eukaryotes. *RNA*, **9**, 644–647.
4. Barrick,J.E., Corbino,K.A., Winkler,W.C., Nahvi,A., Mandal,M., Collins,J., Lee,M., Roth,A., Sudarsan,N., Jona,I. *et al*. (2004) New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc. Natl Acad. Sci. USA*, **101**, 6421–6426.
5. Abreu-Goodger,C., Ontiveros-Palacios,N., Ciria,R. and Merino,E. (2004) Conserved regulatory motifs in bacteria: riboswitches and beyond. *Trends Genet.*, **20**, 475–479.
6. Tatusov,R.L., Koonin,E.V. and Lipman,D.J. (1997) A genomic perspective on protein families. *Science*, **24**, 631–637.
7. Qi,J., Wang,B. and Hao,B.L. (2004) Whole genome prokaryote phylogeny without sequence alignment: a K-string composition approach. *J. Mol. Evol.*, **58**, 1–11.
8. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Ahang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
9. Ciria,R., Abreu-Goodger,C., Morett,E. and Merino,E. (2004) GeConT: gene context analysis. *Bioinformatics*, **20**, 2307–2308.
10. Griffiths-Jones,S., Bateman,A., Marshall,M., Khanna,A. and Eddy,S.R. (2003) Rfam: an RNA family database. *Nucleic Acids Res.*, **31**, 439–441.
11. Kanehisa,M. and Goto,S. (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30.
12. Bengert,P. and Dandekar,T. (2004) Riboswitch finder—a tool for identification of riboswitch RNAs. *Nucleic Acids Res.*, **32**, W154–W159.
13. Bailey,T.L. and Gribskov,M. (1998) Combining evidence using p-values: application to sequence homology searches. *Bioinformatics*, **14**, 48–54.
14. Merino,E. and Yanofsky,C. (2005) Transcription attenuation: a highly conserved regulatory strategy used by bacteria. *Trends Genet.*, **21**, 249–305.